SUBSTITUTE ABSTRACT

I describe several techniques for characterizing molecules based on the shapes of their fields. The minimal distance between two molecular fields is used as a shape-based metric, independent of the underlying chemical structure, and a high-dimensional shape space description of the molecules is generated. I then show how these attributes can be used in creating, characterizing, and searching databases of molecules based on field similarity. In particular, they allow searches of a database in sublinear time. Next, I extend the utility of this approach by describing a way to automatically break molecules into a series of fragments by using an ellipsoidal Gaussian decomposition. Not only can these fragments then be analyzed by the shape metric technique described above, but the parameters of the decomposition themselves can also be used to further organize and search databases. The most immediate application of these techniques is to pharmaceutical drug discovery and design.

IMPLEMENTATION AND EXAMPLES

Specific implementations and examples of the foregoing include the following:

1: Finding The Maximal Overlap (Minimal Field Difference) Between Two Fields A And B

2: Refining The Search Position Via Numerical Or Analytical Derivatives

3: Determining The Shape Space Of A Set Of Molecules

4: Calculating The Position Of A New Structure In A Preconstructed Shape Space

5: Extending The Shape Space

6: Calculating The Maximal Overlap Between A New Structure And A Large, Previously Shape-Space Decomposed Set Of Molecules

7: Using The Shape Space Description To Correlate With Known Biological Activity

8: Examples Of Using The Minimum Field Difference Metric To Organize A Database Of Molecules

9: Examples Of Using The Shape Space Positions To Organize A Database Of Molecules

10: Local Domain Decomposition

11: Constructing An Ellipsoidal Gaussian Representation (EGR).

12: Construction Of Multiple EGR's Containing The Same Number Of EGF's.

13: Constructing Molecular Fragments From An EGR.

14: Evaluating An EGR Fit: The Fragment Adjusted EFF.

15: Construction Of Multiple EGR's With Different Numbers Of EGF's

16: Storage In A Database

17: Comparing Single EGR's To Solve The Atom Assignment Problem

18: EGF Pseudo-Surfaces And Painted EGF's

19: Using A Database Of EGR's For (Molecular/Molecular Fragment) Similarity Evaluation

20: Using EGF's In Similarity Searches

21: Using EGR's To Find A Place In Shape Space; Using Shape Space To Find EGR's

22: Defining An EGR For A Negative Space